



Hassan, Saeed-Ul, Aljohani, Naif R, Idrees, Nimra, Sarwar, Raheem, Nawaz, Raheel ORCID logoORCID: <https://orcid.org/0000-0001-9588-0052>, Martínez-Cámara, Eugenio, Ventura, Sebastián and Herrera, Francisco (2020) Predicting literature's early impact with sentiment analysis in Twitter. Knowledge-Based Systems, 192. p. 105383. ISSN 0950-7051

Downloaded from: <https://e-space.mmu.ac.uk/624666/>

Version: Accepted Version

Publisher: Elsevier BV

DOI: <https://doi.org/10.1016/j.knosys.2019.105383>

Usage rights: Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Please cite the published version

<https://e-space.mmu.ac.uk>

Manuscript Number:

Title: Predicting Literature's Early Impact with Sentiment Analysis in Twitter

Article Type: Full length article

Keywords: Altmetric, Twitter, Sentiment Analysis, User Category, Predicting Citations

Corresponding Author: Dr. Eugenio Martínez Cámara, Ph.D. in Computer Science

Corresponding Author's Institution: University of Granada

First Author: Saeed-Ul Hassan

Order of Authors: Saeed-Ul Hassan; Naif R. Aljohani; Nimra Idrees; Raheem Sarwar; Raheel Nawaz; Eugenio Martínez Cámara, Ph.D. in Computer Science; Sebastián Ventura; Francisco Herrera

Abstract: The traditional bibliometric techniques gauge the research impact through citation-based quantitative indices. However, due to citation lag time, it may take years to address the impact of an article. This paper seeks to measure an early impact of research articles using tweet sentiments associated with them. We claim that the papers cited in positive and neutral tweets have a higher impact than those not cited or cited in negative tweets. Accordingly, we use SentiStrength, and we improve it by incorporating new opinion bearing words of scientific domain in its sentiment lexicons. Then, we classify the sentiment of 6,482,260 tweets linked to 1,083,535 publications covered by Altmetric.com. By using positive and negative tweets as an independent variable and the citation count as the dependent variable, the linear regression analysis shows a weak positive prediction of high citation counts across 16 broad disciplines in Scopus. By introducing an additional indicator, i.e. 'number of unique Twitter users,' the regression model improves the adjusted R-squared value of regression analysis in several disciplines. Overall, the encouraging positive correlation between the tweet sentiments and citations show that Twitter-based opinion may be exploited as a complementary indicator for predicting literature's early impact.

Suggested Reviewers:

Predicting Literature’s Early Impact with Sentiment Analysis in Twitter

Saeed-Ul Hassan^a, Naif R. Aljohani^b, Nimra Idrees^a, Raheem Sarwar^a, Raheel Nawaz^c, Eugenio Martínez-Cámara^{d *}, Sebastián Ventura^{b,e}, Francisco Herrera^{b,d}

^a Information Technology University, 346-B, Ferozepur Road, Lahore, Pakistan

^b Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia.

^c Department of Operations, Technology, Events and Hospitality Management, Manchester Metropolitan University, Manchester, United Kingdom.

^d Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada, 18071 - Granada, Spain

^e Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Córdoba, 14071 - Córdoba, Spain

Abstract

The traditional bibliometric techniques gauge the research impact through citation-based quantitative indices. However, due to citation lag time, it may take years to address the impact of an article. This paper seeks to measure an early impact of research articles using tweet sentiments associated with them. We claim that the papers cited in positive and neutral tweets have a higher impact than those not cited or cited in negative tweets. Accordingly, we use SentiStrenth, and we improve it by incorporating new opinion bearing words of scientific domain in its sentiment lexicons. Then, we classify the sentiment of 6,482,260 tweets linked to 1,083,535 publications covered by Altmetric.com. By using positive and negative tweets as an independent variable and the citation count as the dependent variable, the linear regression analysis shows a weak positive prediction of high citation counts across 16 broad disciplines in Scopus. By introducing an additional indicator, i.e. ‘number of unique Twitter users,’ the regression model improves the adjusted R-squared value of regression analysis in several disciplines. Overall, the encouraging positive correlation between the tweet sentiments and citations show that Twitter-based opinion may be exploited as a complementary indicator for predicting literature’s early impact.

Keywords: Altmetric, Twitter, Sentiment Analysis, User Category, Predicting Citations

* Corresponding author

Email addresses: saeed-ul-hassan@itu.edu.pk (Saeed-Ul Hassan), nraljohani@kau.edu.sa (Naif R. Aljohani), nimraidrees@yahoo.com (Nimra Idrees), raheem.bwl@gmail.com (Raheem Sarwar), r.nawaz@mmu.ac.uk (Raheel Nawaz), emcamara@decsai.ugr.es (Eugenio Martínez-Cámara), sventura@uco.es (Sebastián Ventura), herrera@decsai.ugr.es (Francisco Herrera)

1. Introduction

Altmetric is an umbrella term - many social media platforms, such as Twitter¹, Facebook², CiteULike³ and Mendeley⁴ Readership, can be used as article-level metrics to measure research impact and thus may be referred to by this term (Haustein et al., 2015). With the growth in article-level metric data, the need to provide tools to allow researchers to utilize these data is growing. There are a number of altmetric data aggregators available, including Altmetric.com, ImpactStory and Plum Analytics, to capture article-level web activity and provide the data to researchers. Scholars are increasingly using online platforms to read, bookmark, share, discuss and rate research, which results in a large amount of online data. Mining these data may provide useful insights as an alternative to traditional citation metrics (Priem et al., 2011). Although the popularity of altmetric techniques has been increasing (Nuzzolese et al, 2019), there is a lack of information and evidence on their effectiveness, and the major outstanding challenges are to ensure the use of standards and best practice (Haustein et al., 2015; Bornmann et al, 2019).

The traditional bibliometric methods gauge research impact through citation-based quantitative indices like journal impact factor, h-index, source normalized impact per paper (Haddawy et al, 2017), however, due to citation lag time, a limitation associated with the citation-based quantitative indices, it may take years before the impact of an article can be assessed. With increased usage of the web for scholarly communications, altmetric data (Priem et al., 2011) are enhanced as they capture real-time data from online platforms such as Twitter, Facebook and CiteULike. Therefore, altmetric techniques can be used to measure the early impact of scientific

¹ <https://twitter.com>

² <https://www.facebook.com>

³ <http://www.citeulike.org>

⁴ <https://www.mendeley.com>

literature (Didegah et al, 2018). Twitter is a platform widely used by scholars to share their opinion on research articles (Priem et al., 2011). There is a need to investigate how authentic is this way of measuring impact and whether an article's high tweet count leads to a future high citation count. Altmetric.com captures the tweet count for each research article and also other Twitter demographics, such as the Twitter user category. Understanding the extent to which a tweet linked to a scholarly article conveys opinion about it might help us to understand the importance of the Twitter indicator as a measure of that article's impact.

Consequently, in this paper, we study the influence on the early impact of research literature of the opinions posted on Twitter and other Twitter data. We hypothesize that the papers cited in positive and neutral tweets may have a higher impact than those ones not cited or cited in negative tweets. In order to evaluate that, we measure an early impact of tweet sentiments associated with the research articles disseminated on the Twitter platform using the text of over 6 million tweets - covered by Altmetric.com from July 2010 to June 2016. We explore features, the positive, negative and neutral sentiments, in tweets along with the unique Twitter user. We performed multiple linear regression analysis on our dataset to analyze the use of Twitter as a high or a low predictor of citation count. We first differentiated the counts of negative, positive and neutral tweets from the original altmetric tweet count, then multiple linear regression was applied to tweets, with positive and neutral sentiments as independent variables and the citation count as the dependent variable. To overcome the effect of fake distribution of tweets, unique Twitter user was also introduced into the multiple linear regression analysis as third independent variable to a predict high citation count.

As for the tweet sentiment analysis, we utilize Twitter text messages consist of a maximum of 140 characters⁵ providing the actual text message, also hashtags, usernames, pictures and URLs. We note that Twitter text is often consisting of abbreviations, contractions and acronyms. The text also contains shorten, cut and truncate messages and slang. We found that the lexicon-based sentiment approach, employed by SentiStrength (SentiStrength, 2017) is well suited in our context. Also, a number of studies have shown the effectiveness of SentiStrength over existing tools in analysing the sentiments of tweet text. For instance, Friedrich et al. (2015b) analysed two existing sentiment analysis tools, SentiStrength and Sentiment140, to detect the sentiments in tweets about academic articles. They concluded that by adapting the lexicons of the SentiStrength tool to scholarly terms its efficiency in detecting sentiments in scholarly tweets could be increased greatly. SentiStrength uses an algorithm that simultaneously extracts positive and negative sentiments from short, informal texts (Thelwall, Buckley & Paltoglou, 2012). In this investigation, we propose to improve SentiStrength⁶ program by incorporating new opinion bearing words to update its sentiment lexicon, and to make it more suited for the impact assessment of the tweets on scientific literature.

In order to achieve the objectives of this investigation, our experimental studies are driven by the following research questions:

- What is the influence of the new opinion bearing words in the research domain included in SentiStrength for the impact assessment of tweets on scientific literature?
- What kind of opinions (positive, negative or neutral) do tweets convey about a linked research article?

⁵ Note that the recent advancement on Twitter platform allows up to 280 characters.

⁶ The SentiStrength is freely available for research purposes and its lexicons can be adapted to the appropriate field of interest.

- What is the difference between the disciplines regarding tweets containing positive, negative and neutral sentiments?
- Which Twitter user categories share the most opinions when tweeting about research articles?
- Does a high tweet count with positive sentiments about a research article lead to a future high citation count for that article?
- Can high tweet count with negative sentiments about a research article lead to a future low citation count for that article?

We can summarize the contributions of this study as follows: a) It proposes to use tweet sentiments information associated with the research articles disseminated on the Twitter platform in order to predict the early impact of research articles, b) Keeping the context of research articles and associated tweets in view, we improve SentiStrength – a sentiment analysis program, by incorporating new opinion bearing words to make its lexicon more suited for the impact assessment of tweets on scientific literature and provide its performance comparison against classical SentiStrength, c) further, by employing a multiple linear regression model, we show the relationship between the citation counts and tweet sentiments associated with the research articles, d) finally, our extensive experimental studies show that the tweet sentiments can be used as an indicator for the early impact assessment of scientific research.

The rest of the paper has been organized as follows: Section 2 presents the review on the related Altmetric studies that seeks to measure the impact of Twitter on the dissemination of scientific literature. Section 3 presents the dataset, pre-processing approach to feed the data to SentiStrength model along with the evaluation setup of the classic SentiStrength in comparison with the adopted SentiStrength model. Next, Section 4 provides a detailed discussion of the results. Finally, Section 5 concludes the research findings and highlights future research directions.

2. Literature Review

A number of studies showed that only few tweets about a research article convey much positive or negative sentiment towards the associated article, while most of the tweets appeared to be neutral and sole for information dissemination. Thelwall et al. (2013) performed a pilot study on 270 randomly collected tweets about research articles and analyzed the kinds of opinion the tweets conveyed about the articles and whether the ratio of negative tweets to the overall tweet count might be ignored as a measure of research impact. Their results showed that tweets linked to scholarly articles are mostly objective, as they consist of either the article title or points from a brief summary of the article. Friedrich et al. (2015a) analyzed tweets on the articles and reviews published in Web of Science (WoS) in 2012, captured by Altmetric.com on 487,610 tweets mentioning 192,832 articles. Their results showed that, of all the tweets, 11.0% conveyed positive sentiments and 7.3% negative sentiments, while 81.7% were neutral. Further, the disciplinary analysis showed that tweets from those in disciplines such as Psychology, and the Humanities and Social Sciences generally, convey the highest proportion of sentiments, while those from Physics, Chemistry and Engineering convey the least sentiments about articles in their fields. The study suffered from its small dataset. In contrast, our study performed sentiment analyses on a dataset of over 6 million tweets linked to more than a million research articles captured by Altmetric.com.

Martin Fenner (2013) investigated article-level matrices of PLOS Biology research published in 2010. The study showed that, although some of the highly cited articles had a high number of online viewings, overall there was a low correlation between the numbers of citations and online

views. Thelwall et al. (2013) analyzed the degree of correlation between various altmetric sources and citation counts. The study compared eleven altmetric indicators with Web of Science citations for 76 to 208,739 PubMed articles. The results provided significant evidence of a high altmetric score and high citation count for Twitter, Facebook posts, blogs, research highlights, online media and forums, but very little or no correlation for Google+. While there were insufficient data to support a correlation between the citation count and other indicators, such as LinkedIn, Pinterest, Q&A sites and Reddit.

Costa et al. (2015) performed analysis of various altmetric indicators provided by Altmetric.com and their correlation to the citation count. Their results showed that, though there is a positive correlation between various altmetric indicators and the citation count, the value is very low, showing weak correlation, and they concluded that altmetric indicators do not measure the same impact as traditional methods, such as citation counts. Ravenscroft et al. (2017) investigated the correlation between altmetric score and the research evaluation framework (REF) impact, and their results show that there is little significant correlation.

Houqiang Yu (2017) found that in all the altmetric indicators there is a significant difference between the number of posts (NP) and the number of unique users (NUU). He identified that there is a high to moderate pearson correlation between NP and NUU for various altmetric indicators. He also analyzed the Twitter user count for the various user categories: researcher; practitioner; science communicator; and member of the public. Further, the correlation analysis was also performed on the Twitter user counts in each category and the citation count of the associated research article. The results revealed that, of all the categories, that of researcher

yields the highest correlation value, yet the overall value remains low, similar to the findings in previous studies.

A number of studies have aimed to establish the extent to which Twitter is an authentic measure of the research impact of an academic article; and whether we can use it to predict its citation count (Priem et al., 2011; Thelwall et al., 2013; Costas et al., 2015; Haustein et al., 2015; Eysenbach, 2011; Holmberg & Thelwall, 2014). These studies used simple correlation techniques to find the relationship between the citation count and the raw tweet count. In this direction of research, Konkiel (2016) suggested that a single indicator might not comprehensively measure the impact of research, thus, it would be more beneficial if researchers considered a combination of alternative metrics. Tsou et al. (2015) and Haustein et al. (2016) revealed that a high number of tweets about scholarly articles are from automated bot accounts, and the same accounts tweet hundreds of times about the same article, this can affect Twitter's use as an impact measure. In addition, Yu (2017) identified that there is a considerable disparity between the number of posts (NP) and the number of unique Twitter users, and one possible reason might be multiple activities by a single user about the same article for the sake of self-promotion, advertising or fraud. This, too, makes the altmetric indicator less valuable as a measure of research impact.

In this paper, we explored features: the positive, negative and neutral sentiments in tweets along with the unique Twitter user. We performed multiple linear regression analysis on our dataset to analyze the use of Twitter as a high or a low predictor of citation count. We first differentiated the counts of negative, positive and neutral tweets from the original altmetric tweet count, then

multiple linear regression was applied to tweets, with positive and neutral sentiments as independent variables and the citation count as the dependent variable. To overcome the effect of fake distribution of tweets, unique Twitter user was also introduced into the multiple linear regression analysis as third independent variable to predict high citation count. The results were noticeably improved in disciplines such as Earth and Planetary Sciences, Health Professions and Nursing, Mathematics, and Medicine and Medical Sciences. The result of multiple linear regression analysis, which used negative tweets to predict a low citation count, remained approximately zero across all the disciplines, yet the discipline listed as ‘general’ showed a weak negative prediction.

3. Data and Methodology

This section presents altmetric data and its pre-processing steps along with the employed sentiment analysis approaches that are suited for the opinion mining of Twitter on research articles. Last but not the least, we also show the evaluation results of classic SentiStrength model compared with the proposed model with new opinion bearing words.

3.1 Dataset

The corpus consisted of altmetric data captured by Altmetric.com⁷ from July 2011 to June 2016. In total, 1,083,535 research articles having at least one tweet and one citation (till Feb 2017) were extracted from the altmetric dataset. Noted that Altmetric.com assigns a unique ID, ‘altmetric_id’, to each article and captures it for every tweet about that article. The text from tweets associated with tweet_ids was scraped from Twitter.com by passing tweet IDs through

⁷ The data was received from the Altmetrics.com in JSON file format. As per the agreement, the author can not publically disseminate the copy of this data. However, the same data may freely be obtained by the scientific community from the Altmetric.com for research purpose.

Twitter API. Further, the text of 6,482,260 tweets was scraped from Twitter.com by passing tweet ids provided by Almetric.com methods through Twitter's API. Against each article, Almetric.com records the Unique User Count (UUC) of those who tweet about it, and this information was extracted from the altmetric dataset. User information is captured in altmetric data, categorizing users as: researcher; practitioner; science communicator; or member of the public. These categories are assigned on the basis of the information in the user's profile, the types of journals that they are linked to and their 'friends' lists (Altmetric LLP, 2017). Finally, the citation count of research articles was collected from Scopus API and disciplinary information assigned to each article as per the Scopus subject-category scheme employed by Haddawy et al. (2017).

3.2 Sentiment analysis approaches

As mentioned in introduction, we hypothesize that papers cited in positive and neutral tweets have a higher impact than those ones not cited or cited in negative tweets. In general, the sentiment analysis can be applied at three levels of detail: (1) document level; (2) sentence level; and (3) entity/aspect level (Liu, 2012). Our evaluation is based on document/tweet level. The approach utilized in this study is a lexicon-based method grounded in counting the number of words in a given text that are in a sentiment lexicon (SentiStreth, 2017). Here the opinion/sentiment is acquired by a lexicon-based sentiment classifier using a sentiment lexicon and a combination of sentiment word scores and query term–sentiment word proximity scores.

Our study utilized Twitter text messages consist of a maximum of 140 characters⁸ providing the actual text message, also hashtags, usernames, pictures and URLs. Note that Twitter text is often consisting of abbreviations, contractions and acronyms. The text also contains shorten, cut and truncate messages and slang. Therefore, we found that the lexicon-based sentiment approach, employed by SentiStrength is well suited in our context. Also, a number of studies have shown the effectiveness of SentiStrength over existing tools in analysing the sentiments of tweet text. For instance, Friedrich et al. (2015b) analysed two existing sentiment analysis tools, SentiStrength and Sentiment140, to detect the sentiments in tweets about academic articles. They concluded that by adapting the lexicons of the SentiStrength tool to scholarly terms its efficiency in detecting sentiments in scholarly tweets could be increased greatly. SentiStrength uses an algorithm that simultaneously extracts positive and negative sentiments from short, informal texts (Thelwall, Buckley & Paltoglou, 2012).⁹

Following the work of Liu et al. (2017), we identified a list of the most commonly used terms in tweets about a research article that were either to praise it or to convey negative sentiments towards the published work. Positive tweets about such an article usually contain words or phrases such as a ‘compelling article’, ‘fundamental study’, ‘remarkable finding’ or ‘novel technique’, and negative tweets contain words or phrases such as a ‘biased article’, ‘bad idea’, ‘fake’, ‘fallacy’ or ‘weak conclusion’. All such terms were searched for in the tweet dataset, and the resultant tweets were considered carefully. Further, the terms that were found in many tweets, as either positive or negative opinion, were added to the SentiStrength lexicons, and around 80 positive and negative terms were added in this way. In addition to these terms, we

⁸ Note that the recent advancement on Twitter platform allows up to 280 characters.

⁹ The SentiStrength is freely available for research purposes and its lexicons can be adapted to the appropriate field of interest.

searched the SentiStrength lexicons for the words found in the tweets dataset one by one, and the tweets containing such words were intellectually analyzed. Examining the lexicons, we noted that many terms used in tweets, such as ‘death’, ‘war’, ‘accident’, ‘germs’ and ‘care’, were science-specific terms and had not been employed by the authors to convey opinion about the linked article. These terms were causing the assignment of false positive or negative sentiment to tweets, therefore we decided to remove them from the SentiStrength lexicons. In total, 148 such terms were removed.

We found that tweets sometimes contain research-specific terms taken from article’s title, and do not convey an actual opinion on the article. By comparing each word in a tweet text string with each word in the linked article’s title string, then removing the word from the tweet’s text, matched terms such as ‘cancer’, ‘disaster’ and ‘harm’ were removed to prevent any false assignment of sentiments, and this greatly improved SentiStrength’s efficiency in detecting sentiment. Note that Friedrich et al. (2015b) also adopted this practice in their work on analyzing tweet sentiments. Further, URLs, # signs and user mentions (@username) were considered not to carry any opinion about the article, and were duly removed from tweets’ text to avoid any false assignment of sentiment. Moreover, the language of each tweet’s text was detected using the R programming language and those tweets in a language other than English were filtered out, as SentiStrength lexicons are English based and inclusion of tweets in other languages might have resulted in the assignment of false sentiments to such tweets. Finally, using SentiStrength with our adapted lexicons, the sentiments of the remaining 5,341,800 tweets were detected. The data pertaining to the adopted lexicons can be found in Appendix A, Table A-1 through Table A-3.

Finally, to provide the sentiment of tweet text, our adopted SentiStrength model could provide two sentiment-strength values as output: -1 (not negative) to -5 (extremely negative); and 1 (not positive) to 5 (extremely positive). Further, the tweets for which sentiment strength was detected to be between 2 to 5 were counted as positive tweets, and those for which sentiment strength was detected to be between -2 to -5 were counted as negative. The rest, for which sentiment strength was between 1 and -1, were counted as neutral tweets. In this way, the count of positive, negative and neutral tweets was identified.

Table 1: Evaluation of classification models

Models	Precision	Recall	F1 Score	Accuracy
New Lexicon	0.660	0.355	0.215	0.570
SentiStrength	0.569	0.489	0.476	0.659
SentiStrength + New Lexicon	0.642	0.581	0.576	0.721

3.3 Evaluation of SentiStrength Models

In this section we present the evaluation of classification models. For the evaluation of models, we annotate a subset of tweets from the original dataset. This subset contains 2544 tweets on English language on the publications belonging to various disciplines such as Biomedical & Health Sciences; Life & Earth Sciences; Mathematics & Computer Science; Physical Sciences & Engineering; and Social Sciences & Humanities. Further, we annotate these tweets into positive, negative or neutral sentiments by keeping the context of the papers in view with the help of two annotators. We get the label dataset of selected tweets with a reasonable 86% inter-annotator agreement classified into positive (34.7%), negative (35.7%) and neutral (29.6%) classes. Finally, we run the following three models a) New Lexicon, b) classic SentiStrength, and c)

SentiStrength + New Lexicon on the selected tweets and compare the results with that of humanly annotated labels.

Table 1 shows the evaluation results of SentiStrength and New Lexicon models compared to our adopted SentiStrength + New Lexicon model. We find that our adopted SentiStrength + New Lexicon model gives high accuracy in predicting tweet sentiments with average accuracy of 72.1%, compared to SentiStrength and New Lexicon models that show 65.9% and 57% average accuracies, respectively. The adopted SentiStrength + New Lexicon model also shows high F1 score and Recall compared to SentiStrength and New Lexicon models. Interestingly, the New Lexicon model show the highest avg. precision, but, this is at the cost of very low avg. recall. Overall, the aim of this evaluation is to show the suitability of combining the words of SentiStrength and the words of New Lexicon. We show that the new words of SentiStrength do not hinder the set of new words. Instead, the combination of the two lexicons, SentiStrength and the new one allow to improve the recall score at the cost of precision from 66% to 64.2% - that makes the F1 score higher than using independently the two lexicons. Thus, the combination of new words with the lexicon of classic SentiStrength appears to be well suited for the classification task of tweet sentiments.

4. Analyses and Results

This section presents analysis and discussion on our results. Using the adopted SentiStrength with new lexicons, we show the distribution of tweet sentiments across broad disciplines ranging from Agricultural, Biological Sciences & Veterinary through Social Sciences and across Twitter user categories. Finally, we discuss regression analysis between the tweet sentiments and citations.

4.1 Distribution of tweet sentiments across scholarly disciplines

Of the total 5,341,800 tweets, 77.9% were found to be neutral, 14% positive and 10.3% negative. We found some tweets were labelled under more than one class, therefore, the sum of all classes is slightly above 100%. The cross-disciplinary analysis of the sentiments of these tweets showed that the majority of tweets are neutral in all disciplines, and that the percentage of positive sentiments is slightly higher than that of negative sentiments in most disciplines (see Fig. 1), which confirmed previous findings (Thelwall et al., 2013; Friedrich, et al., 2015b). We found that the field of Economics, Business and Decision Sciences has a low tweet count of 67,946, yet it has the highest count of tweets conveying positive sentiments (17.5%), and that, in the disciplines of both 'General' and Health Professions and Nursing, tweets also have a high percentage of positive and negative sentiments, and it can be concluded that their tweets linked to articles in their fields convey more sentiments than in other disciplines.

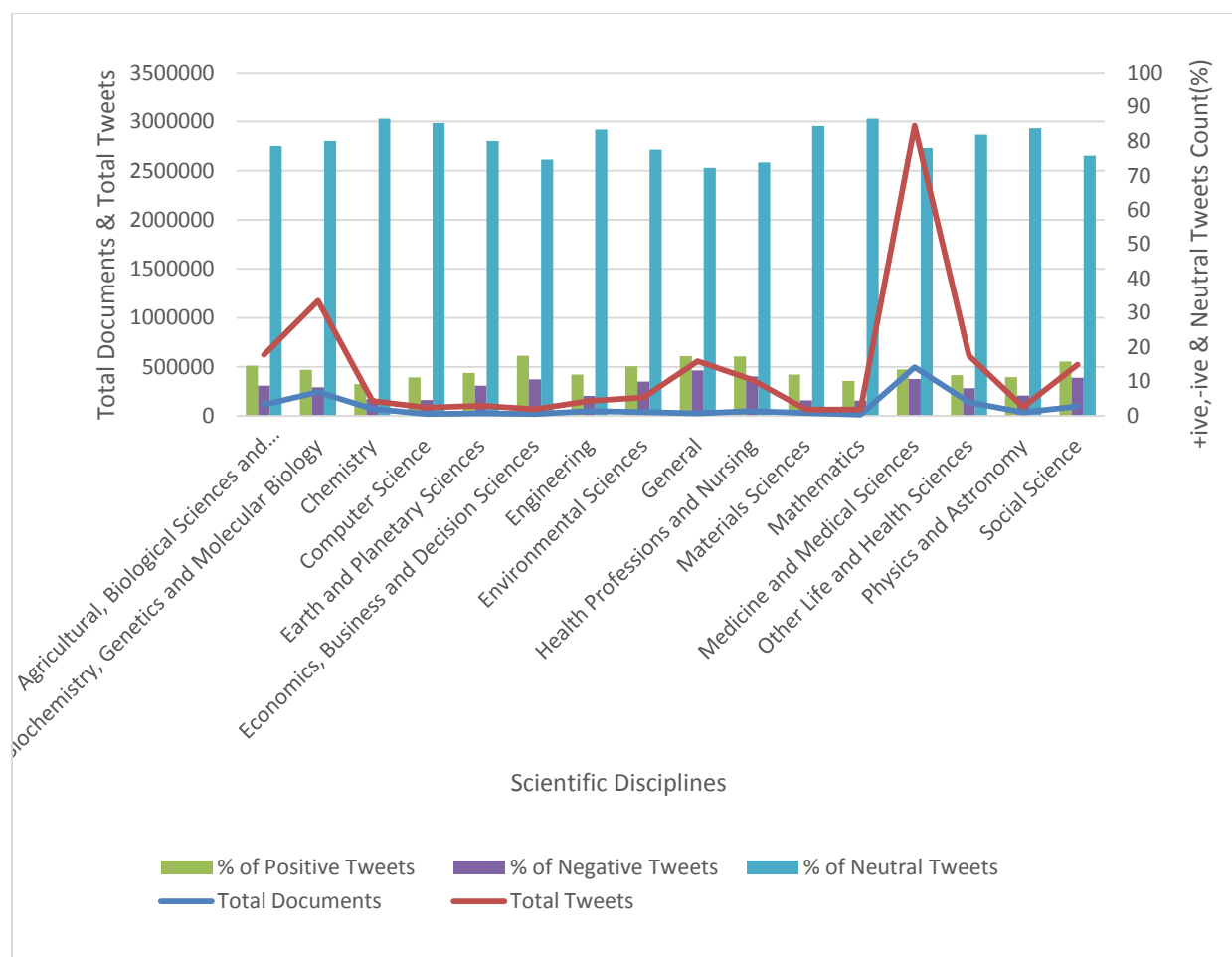


Figure 1. Distribution of tweets with positive, negative or neutral sentiments across scholarly disciplines

4.2 Distribution of tweet sentiments among Twitter user categories

The number of articles and number of positive, negative and neutral tweets were counted in each of the four user categories in which an article has at least one Twitter user interaction. The data were analyzed against these user categories, and the results showed that although the total tweet count is higher for the category ‘member of the public’, those in the other three categories (researcher, practitioner and science communicator) convey a higher number of positive and negative sentiments in their tweets (see Fig. 2). This comparatively high percentage of sentiments in the latter three categories is as expected, because these users interact and use

research more in their daily routine than general members of the public, thus are more likely to convey sentiments and opinion in their tweets.

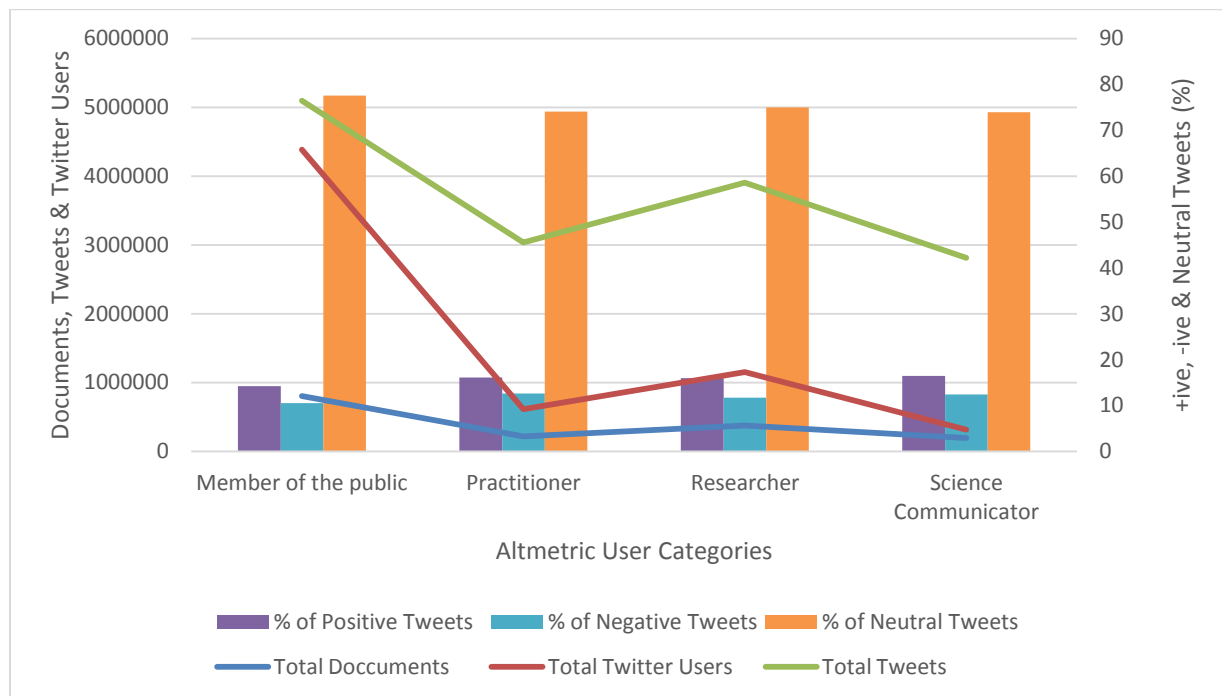


Figure 2. Distribution of tweets with positive, negative or neutral sentiments across user categories

4.2 Multiple linear regression analysis on Twitter data

Previous studies have used raw tweet counts of research articles to analyze the correlation of tweet count to the citation count without considering whether the opinion about the article in the tweets is positive or negative (Priem et al., 2011; Thelwall et al., 2013; Costas et al., 2015; Haustein et al., 2015; Hassan et al., 2017). In this study, we removed from the total tweet count all those that convey negative sentiments to the article, then performed multiple linear regression analysis on the tweet dataset to predict the citation count for the article. We first applied linear regression to the positive and neutral tweet counts as independent variables and to the citation count as dependent variable. The R-squared and adjusted R-squared values of multiple linear

regression analysis remained low, which showed a weak prediction of citation count. As a single Twitter user can send several tweets about an article, to reduce the effect of fake distribution we next introduced a third variable, UUC, to the multiple linear regression model. Of the variables that we used as independent variables in our regression model, as shown in Table 2 the P value is approximately zero. This makes it a significant variable.

Introducing the UUC variable to the multiple linear regression analysis slightly improved the results, and the adjusted R-squared value of regression analysis noticeably improved in disciplines such as Earth and Planetary Sciences, Health Professions and Nursing, Mathematics, and Medicine and Medical Sciences (see Fig. 3). We presented adjusted R-squared values throughout the analysis due to the fact that R-squared increases every time we add a new variable in the model however adjusted R-squared value increases only if the new variable improves the model.

Table 2: Coefficients of Regression Model with Positive and Neutral Tweets

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	12.761125	0.038188	334.16	<2e-16 ***
Positive tweets	0.254208	0.011579	21.95	<2e-16 ***
Neutral tweets	0.337588	0.004511	74.83	<2e-16 ***
UUC	0.035927	0.001179	30.48	<2e-16 ***

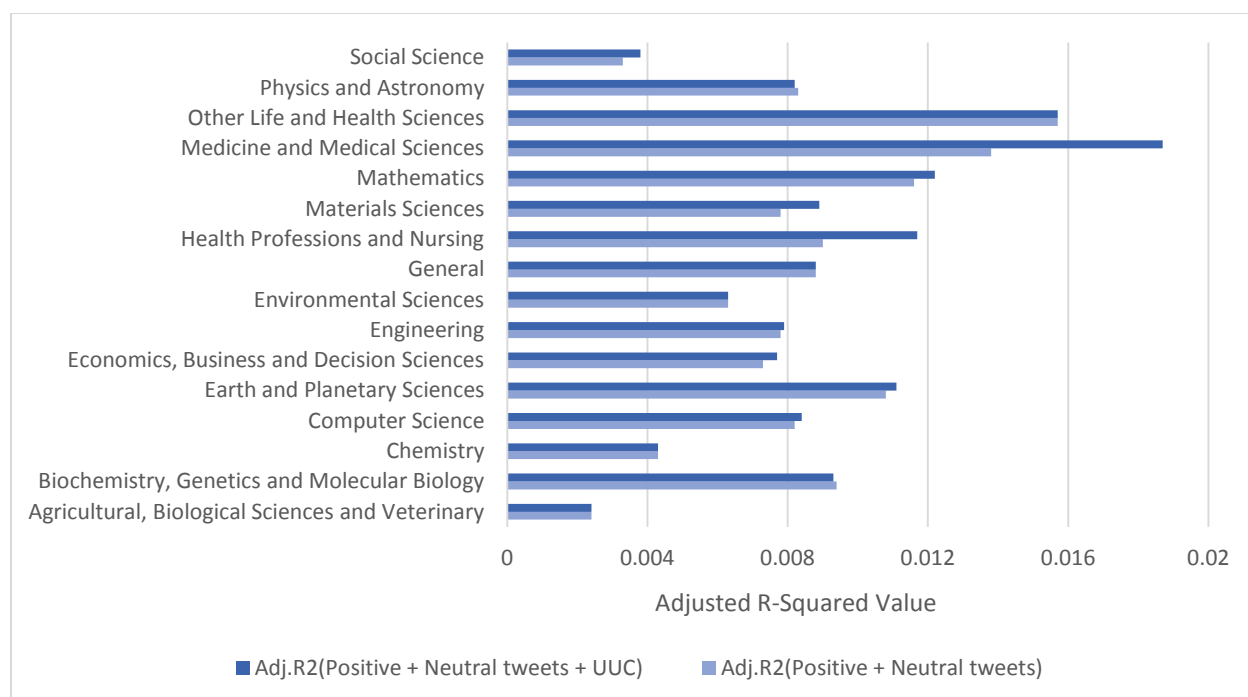


Figure 3. Adjusted R-squared value by using independent variables ‘Positive + Neutral tweets’ and ‘Positive + Neutral tweets + UUC’ across disciplines

To analyze whether tweets with negative sentiments can be used as an early indicator of a future low citation count, linear regression was applied to the negative tweet count as an independent variable and to the citation count as dependent variable, and the adjusted R-squared value of regression analysis remained low for all disciplines; only the discipline of ‘General’ showed a weak negative prediction. The variable of UUC was used as a second independent variable P value for the negative tweet count, and the UUC was approximately zero, which makes these variables significant for the regression model (see Table 3).

Table 3: Coefficients of Regression Model with Negative Tweets

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	13.867750	0.035818	387.18	<2e-16 ***
Negative tweets	0.455607	0.011987	38.01	<2e-16 ***
UUC	0.078219	0.001071	73.03	<2e-16 ***

Introducing UUC to multiple linear regression analysis on negative tweets increased the adjusted R-squared value in disciplines such as Medicine and Medical Sciences, Health Professions and Nursing, and Material Science (see Fig. 4). Nevertheless, the increase in values is simply the effect of UUC, which supports the theory that if there are large numbers of users performing altmetric activity on a scholarly article, the article must be popular and, most likely, will receive more citations in future. In addition, a good line for future work is the study of the influence of negative tweets on citations - to explore if the negative tweets may hinder the future citations.

Further, we performed comparative analysis between the citation count and the adjusted R-squared value using both positive and negative tweets (see Fig. 5). The adjusted R-squared values were higher when positive tweet parameters were used than negative, as expected. This is because, ideally, when the count of positive tweets about an article is high then that article's citation count is also high.

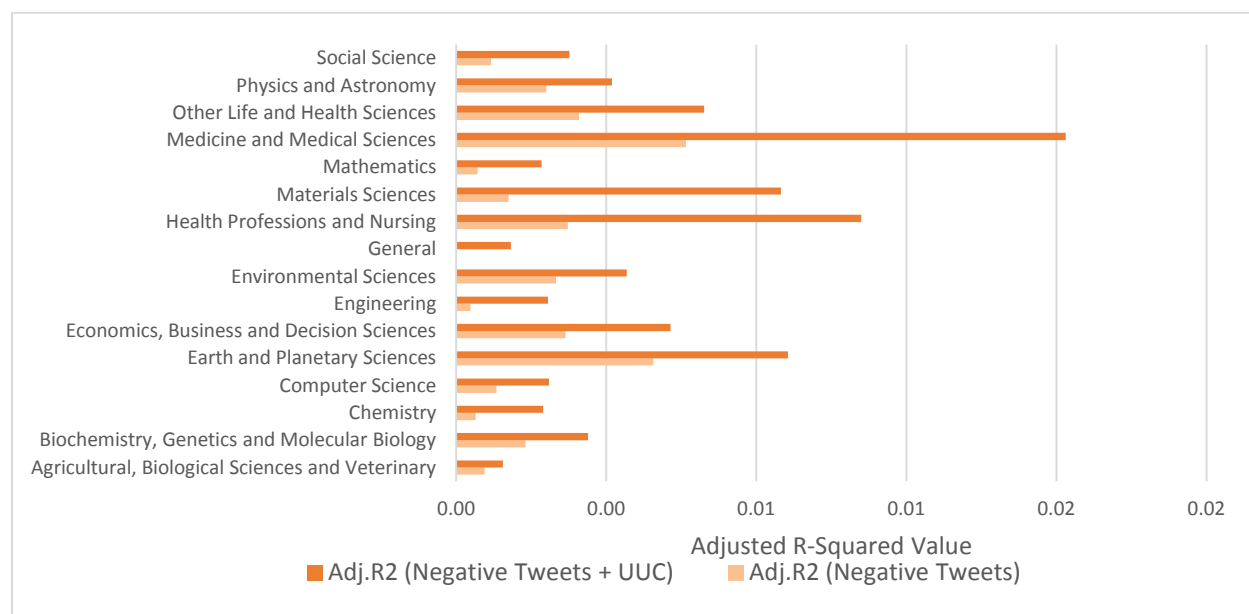


Figure 4. Adjusted R-squared value by using independent variables ‘Negative tweets’ and ‘Negative tweets + UUC’ across disciplines

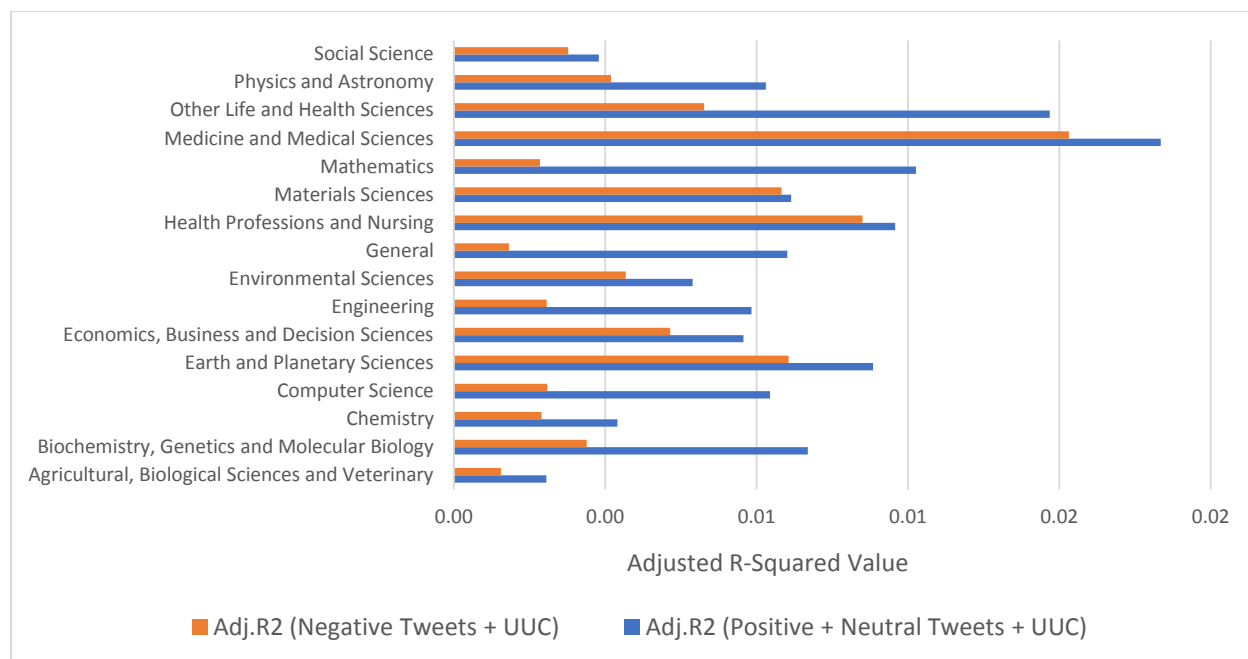


Figure 5. Positive and negative tweets for adjusted R-squared value distributed across disciplines

5. Concluding Remarks

The traditional bibliometric techniques gauge the research impact through citation-based quantitative indices, such as, journal impact factor and h-index. However, due to citation lag time, a limitation associated with citation-based quantitative indices, it may take years before the impact of an article can be assessed. This study proposed to measure an early impact of tweet sentiments associated with the research articles disseminated on the Twitter. In order to show that, at first, we improve SentiStrength – a sentiment analysis system, by incorporating new opinion bearing words to update the lexicon more suited for the impact assessment of tweets on scientific literature. We show that the new opinion bearing words in the research domain included in SentiStrength for the impact assessment of tweets on scientific literature help improve the predicting power of classic SentiStrength model. Thus, the techniques employed can be further be exploited for the assessment of tweets pertaining to the scientific literature. Further,

to evaluate the use of Twitter as an altmetric means of gauging the early impact of a research article', sentiment analysis is performed using tweets about scientific articles indexed in Altmetric.com from July 2011 to June 2016. We found that the papers cited in positive and neutral tweets have a higher impact than those ones not cited or cited in negative tweets. Across the fields such as Economics, Business and Decision Sciences, Health Professions and Nursing, and 'General', tweets convey a comparatively high percentage of sentiments; whereas the field of Chemistry conveys the least. Furthermore, across the twitter-user categories tweet counts are lower among the user categories of researcher, practitioner and science communicator than that of 'member of the public', but that these categories convey more sentiments in their tweets.

One of the limitations of this study lies in the aggregation of sentiment counts across user categories. In Altmetric.com data, a tweet_id can be assigned to multiple user categories. Since Altmetric.com stores Twitter demographics at document level rather than at tweet level, there is no straightforward way to establish how many tweets have actually been sent by a specific category of users. Future studies may consider looking at ways to differentiate this count to achieve superior analysis at user-category level. Another limitation of our study lies in the use of the category "member of the public" of Twitter users. Note that Altmetric.com assigns Twitter users to the categories "researcher", "practitioner", and "science communicator". All other users are assigned to the category "member of the public". In other words, this is a catch-all category to assign some category to the Twitter users where Altmetric.com was unable to assign a proper category to. Therefore, results based on the category "Member of the public" are less useful than results based on other categories.

In future work, instead of assigning its sentiment to three categories (positive, negative and neutral) we seek to establish a tweet's strength of sentiment. Thus, tweets with higher positive strength will be assigned a higher weight in evaluating an article's research impact, or other opinion distribution models can be exploited (Kim et al., 2018; Qiu et al., 2019) – to improve citation prediction. It is possible that, while recent articles may receive much attention online since internet usage by scholars has recently increased, due to the short interval since publication they may have a low citation count. Therefore, to improve the results we need to consider the timespan since publication when predicting citation counts through regression analysis using tweet sentiments. Also meta-knowledge (Thompson et al., 2017; Shardlow et al., 2018) and discourse context (Ananiadou et al, 2013) of tweet text can be exploited using state-of-the-art natural language (Batista-Navarro et al, 2013) and deep learning models (Jahangir et al., 2017) to better understand the impact of tweets. In addition, we recommend to be cautious in the assertion related to a higher number of tweets about a paper increase its likelihood of citation. If most of the tweets are negative, it is likely that the paper will not be cited. We think that the study of the influence of negative tweets may be a nice future work. The future studies may also include non-English tweets in experimental dataset for better coverage of tweet sentiments in altmetric data. Last but not the least, the correlation between the Twitter-based sentiments and citations show that the relationship between these indices are encouraging and may be used as a complementary indicator for predicting literature's early impact, however, further investigation is desired.

Acknowledgments

We would like to thank Altmetric.com for granting us the access to their dataset for research purposes. This work was partially supported by the Spanish Ministry of Science and Technology under the projects TIN2017-89517-P and TIN2017-83445-P, and a grant from the Fondo

Europeo de Desarrollo Regional (FEDER). Eugenio Martnez Camara was supported by the Juan de la Cierva Formación Programme (FJCI-2016-28353) from the Spanish Government.

References

- Altmetric LLP (2017), Fetching detailed article level metrics for an article
<https://help.altmetric.com/support/solutions/articles/6000086844-sample-api-response>,
Retrieve on Dec 19, 2017.
- Ananiadou, S., Thompson, P., & Nawaz, R. (2013, March). Enhancing search: Events and their discourse context. In International Conference on Intelligent Text Processing and Computational Linguistics (pp. 318-334). Springer, Berlin, Heidelberg.
- Batista-Navarro, R. T., Kontonatsios, G., Mihăilă, C., Thompson, P., Rak, R., Nawaz, R., ... & Ananiadou, S. (2013, March). Facilitating the analysis of discourse phenomena in an interoperable NLP platform. In International Conference on Intelligent Text Processing and Computational Linguistics (pp. 559-571). Springer, Berlin, Heidelberg.
- Bornmann, L., Haunschild, R., & Adams, J. (2019). Do altmetrics assess societal impact in a comparable way to case studies? An empirical test of the convergent validity of altmetrics based on data from the UK research excellence framework (REF). *Journal of Informetrics*, 13(1), 325-340.
- Costas, R., Zahedi, Z., & Wouters, P. (2015). Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, 66(10), 2003-2019.
- Didegah, F., Bowman, T. D., & Holmberg, K. (2018). On the differences between citations and altmetrics: An investigation of factors driving altmetrics versus citations for finnish articles. *Journal of the Association for information Science and Technology*, 69(6), 832-843.
- Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of medical Internet research*, 13(4).
- Fenner, M. (2013). What can article-level metrics do for you?. *PLoS biology*, 11(10), e1001687.
- Friedrich, N., Bowman, T. D., & Haustein, S. (2015a). Do tweets to scientific articles contain positive or negative sentiments. <http://altmetrics.org/altmetrics15/friedrich/>, Retrieve on Jan 05, 2017.
- Friedrich, N., Bowman, T. D., Stock, W. G., & Haustein, S. (2015b). Adapting sentiment analysis for tweets linking to scientific papers. *arXiv preprint arXiv:1507.01967*.
- Haddawy, P., Hassan, S. U., Abbey, C. W., & Lee, I. B. (2017). Uncovering fine-grained research excellence: The global research benchmarking system. *Journal of Informetrics*, 11(2), 389-406.

- Hassan, S. U., Imran, M., Gillani, U., Aljohani, N. R., Bowman, T. D., & Didegah, F. (2017). Measuring social media activity of scientific literature: an exhaustive comparison of scopus and novel altmetrics big data. *Scientometrics*, 113(2), 1037-1057.
- Haustein, S., Bowman, T. D., Holmberg, K., Tsou, A., Sugimoto, C. R., & Larivière, V. (2016). Tweets as impact indicators: Examining the implications of automated “bot” accounts on Twitter. *Journal of the Association for Information Science and Technology*, 67(1), 232-238.
- Haustein, S., Costas, R., & Larivière, V. (2015). Characterizing social media metrics of scholarly papers: The effect of document properties and collaboration patterns. *PloS one*, 10(3), e0120495.
- Holmberg, K., & Thelwall, M. (2014). Disciplinary differences in Twitter scholarly communication. *Scientometrics*, 101(2), 1027-1042.
- Jahangir, M., Afzal, H., Ahmed, M., Khurshid, K., & Nawaz, R. (2017, September). An expert system for diabetes prediction using auto tuned multi-layer perceptron. In 2017 Intelligent Systems Conference (IntelliSys) (pp. 722-728). IEEE.
- Kim, H. J., Lee, J., Chae, D. K., & Kim, S. W. (2018). Crowdsourced promotions in doubt: Analyzing effective crowdsourced promotions. *Information Sciences*, 432, 185-198.
- Konkiel, S. (2016). Altmetrics: diversifying the understanding of influential scholarship. *Palgrave Communications*, 2, 16057.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.
- Liu, X. Z., & Fang, H. (2017). What we can learn from tweets linking to research papers. *Scientometrics*, 111(1), 349-369.
- Nuzzolese, Andrea Giovanni, Paolo Ciancarini, Aldo Gangemi, Silvio Peroni, Francesco Poggi, and Valentina Presutti. (2019). Do altmetrics work for assessing research quality?." *Scientometrics*, 119 (2), 539–562
- Priem, J., Piwowar, H., & Hemminger, B. (2011). Altmetrics in the wild: An exploratory study of impact metrics based on social media. In *Metrics 2011: Symposium on Informetric and Scientometric Research*. New Orleans, USA.
- Qiu, J., Lin, Z., & Shuai, Q. (2019). Investigating the opinions distribution in the controversy on social media. *Information Sciences*, 489, 274-288.
- Ravenscroft, J., Liakata, M., Clare, A., & Duma, D. (2017). Measuring scientific impact beyond academia: An assessment of existing impact metrics and proposed improvements. *PloS one* 12.3, e0173152
- SentiStrength (2017), SentiStrength <http://sentistrength.wlv.ac.uk/>, Retrieve on March 10, 2017.
- Shardlow, M., Batista-Navarro, R., Thompson, P., Nawaz, R., McNaught, J., & Ananiadou, S. (2018). Identification of research hypotheses and new knowledge from scientific literature. *BMC medical informatics and decision making*, 18(1), 46.

- Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1), 163-173.
- Thelwall, M., Tsou, A., Weingart, S., Holmberg, K., & Haustein, S. (2013). Tweeting links to academic articles. *Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics*, (17), 1-8.
- Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PloS one*, 8(5), e64841.
- Tsou, A., Bowman, T. D., Ghazinejad, A., & Sugimoto, C. R. (2015, June). Who tweets about science?. In *ISSI*.
- Thompson, P., Nawaz, R., McNaught, J., & Ananiadou, S. (2017). Enriching news events with meta-knowledge information. *Language Resources and Evaluation*, 51(2), 409-438.
- Yu, H. (2017). Context of altmetrics data matters: an investigation of count type and user category. *Scientometrics*, 111(1), 267-283.

Appendix A

Table A1 lists all the terms that were added to the existing SentiStrength lexicon file (EmotionLookupTable.txt). Table A2 lists the adapted idioms in SentiStrength lexicons file (IdiomLookupTable.txt), Table A3 lists all the terms that were used as scientific terminologies and were causing false positive or false negative, and thus were removed from the SentiStrength lexicons file (EmotionLookupTable.txt). The original version of SentiStrength lexicons can be downloaded from: http://sentistrength.wlv.ac.uk/SentStrength_Data_Sept2011.zip.

Table A 1. Terms added to SentiStrength lexicon (EmotionLookupTable.txt)

Terms with a positive sentiment				
sober	soberness	fascinating	clearest (clear* ¹⁰)	fundamental
novel	novelties	fundamentalness	brac* (bracing, brace)	neat study
ground-breaking	novelness	fascinatingly	sound* (sound, sounding)	groundbreaking, ground breaking
worthy	fundament	fundamentally	astonish* (astonish, astonishing)	believe
big*	watershed	unprecedented	landmark	worthful
clever	astound* (astound, astounding)	comprehensive	serious (sentiment updated from negative to positive)	leap study
soberly	compelling	neat research	great systematic	stunning
novelly	remarkable	leap research	incredible	sobering
elegant	productive	extraordinary	fundamentalist	intriguing
Terms with a negative sentiment				
misreporting	fatuous	biased, bias	flaws	not right
joke (sentiment updated from positive to negative)		pilgrims, plagiarized		

¹⁰ Some of the terms are ending with a wild card * which means It can be any word starting with that term for example term astonish* mean both ‘astonish’ and ‘astonishing’.

Table A 2. Adapted idioms in SentiStrength lexicon (IdiomLookupTable.txt)

Idiom Lookup			
wat up	new evidence	how are you	ground-breaking
new way	new research	what's good	less scientific
whats up	shock horror	game changer	new meta analysis
wuts good	breaking news	new analysis	thought provoking
new study	worth reading	felt compelled	thought-provoking
what's up	feel compelled	whats good	ground breaking
it hanging			

Table A 3. Terms removed from SentiStrength lexicon (EmotionLookupTable.txt)

Emotional Lookup			
bug	pains	confes*	prohibit*
war	rape*	corrupt	sufferer*
fat	shark	decease	suffering
foe	tears	default	hazardous
gay	fatty	disease	incurable
gun	fears	dispute	injurious
ill	fever	invade*	corruption
baby	flame	leakage	partition*
bomb	germs	leaking	slaughter*
burn	grave	molest*	unemployed
bury	abrupt	paining	catastrophe
care	absent	poison*	elimination
clog	addict	pollut*	emergencies
cold	afraid	poverty	incompatib*
dead	attack	prison*	compel (changed from negative to positive)
deny	babies	rapist*	dizzy
drag	bother	suffers	enemy
drop	brutal	suicide	jail*
duty	burden	terror*	kick*
envy	cancel	thirsty	adverse
leak	cancer	victim*	against
pain	charge	weapon*	capture
shy*	costly	fatigue	collide
evil	danger	illegal	Emotional

Table A 3. Terms removed from SentiStrength lexicon (EmotionLookupTable.txt) – continued

Emotional Lookup			
fear	injury	illness	outbreak*
feud	mourn*	jobless	paralysis
fist	pained	accident	paralyzed
germ	painf*	cannibal	hunter
grab	raping	casualty	arbitrary
hazy	severe	collapse	crime
homo	spill*	comfort*	death
alarm	suffer	contrary	decay
alien	terror	criminal	devil
argue	thirst	decrease	infect
avoid	feared	disorder	injure
blunt	fierce	pressur*	abandon
bribe	fought	suffered	absence
broke	gunmen	symptom*	collision
cared	hazard	terrori*	dizziness
chase	hunger	haziness	eliminate
choke	hungr*	abnormal*	emergency
cramp			

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: